REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-03-

0038

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. R.
		01 MAY 99 TO 30 NOV 01

4. TITLE AND SUBTITLE
"A COMPETITIVE-MARKET APPROACH TO DISTRIBUTED RESOURCE ALLOCATION"

5. FUNDING NUMBERS
F49620-99-1-0264

6. AUTHOR(S)
Douglas S. Reeves

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
North Carolina State University
Computer Science
Raleigh, NC 27695-7514

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
4015 Wilson Blvd, Room 713
Arlington, VA 22203-1954

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
F49620-99-1-0264

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE, DISTRIBUTION UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Resource allocation is an issue in any system for which resources are in short supply. In computer networks, the resources in question are bandwidth, buffer space, and processing time, and they are allocated at multiple scales of time and quantity (e.g. from a single user's packets over a 30 second interval to a network providers backbone capacity for a I year interval). For military networks and computers, these resources may be scarce or limited during attack, at remote sites, etc. Decisions about resource allocation should be made in accordance with some overall policy. In many instances, this policy is referred to as the "fairness" in making resource allocation decisions. A policy that is reasonable in one situation (e.g., "normally, everyone shares equally the available bandwidth') may be unsuitable in another situation (e.g., "in times of emergency, high priority tasks get their bandwidth requests satisfied before low priority tasks are considered"). In this report, we summarize the progress we have made in applying resource pricing principles to network and computer resource allocation. In each section we explain what we did, and why it is useful, followed by a list of the "outputs" of that effort. We conclude by summarizing the major findings, and describing some future directions and open problems.

20030227 180

14. SUBJECT TERMS

15. NUMBER OF PAGES
12

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

FINAL REPORT

"A Competitive-Market Approach to Distributed Resource Allocation"

Douglas S. Reeves

AFOSR Contract F49620-99-1-0264

April 1999 to April 2002

December, 2002

Problem Statement and Motivation

Resource allocation is an issue in any system for which resources are in short supply. In computer networks, the resources in question are bandwidth, buffer space, and processing time, and they are allocated at multiple scales of time and quantity (e.g. from a single user's packets over a 30 second interval to a network provider's backbone capacity for a 1 year interval). For military networks and computers, these resources may be scarce or limited during attack, at remote sites, etc.

Decisions about resource allocation should be made in accordance with some overall policy. In many instances, this policy is referred to as the "fairness" in making resource allocation decisions. A policy that is reasonable in one situation (e.g., "normally, everyone shares equally the available bandwidth") may be unsuitable in another situation (e.g., "in times of emergency, high priority tasks get their bandwidth requests satisfied before low priority tasks are considered").

For military applications, resource allocation decision-making must be responsive to changing conditions, must be flexible enough to support many types of policies and priorities, and must provide incentives for "responsible use" and disincentives for misuse or attack. We have developed a method of resource allocation that has these qualities,

based on the notion of *resource pricing*. The goal of pricing is to automatically implement the allocation policy determined by the resource owner. The allocation is determined by the user's ability to pay for the resource. The "wealth" of a user may be determined in many ways, based for instance on security level, importance of function, etc.

In this report, we summarize the progress we have made in applying resource pricing principles to network and computer resource allocation. In each section we explain what we did, and why it is useful, followed by a list of the "outputs" of that effort. We conclude by summarizing the major findings, and describing some future directions and open problems.

Topic I: Competitive Bandwidth Pricing for Congestion Control

Goal

Develop a method of resource allocation that is flexible, adapts well to rapidly changing resource conditions and user demands, and is reasonably simple and efficient to implement. The method should support a variety of policies. During periods of congestion or resource scarcity, allocation based on a variety of goals can be easily accomplished, with more important uses given priority over less important uses.

Approach

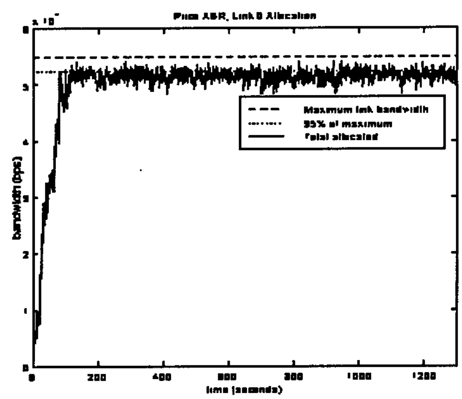
We proposed the use of *congestion pricing*. We showed that it works extremely well for rate allocation in ATM networks. It supports a variety of allocation policies (Pareto optimality, max-min fair, proportional fair) with minor modifications, and is the most flexible method we know for this purpose. It has been implemented and demonstrated on real traffic traces. The method should work well for any rate-based resource allocation.

Results

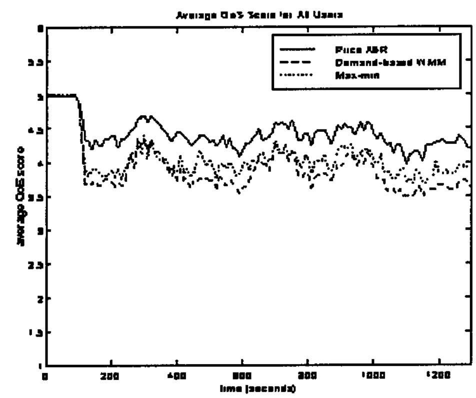
- Errin Fulp graduated in September 1999 with a PhD in Computer Engineering. His thesis title was "Resource Allocation and Pricing for QoS Management in Computer Networks". After a 9-month post-doc at NC State, he joined Wake Forest University as an Assistant Professor of Computer Science.
- We presented a paper at DISCEX II (Defense Information Security Conference and Expo) in Anaheim, CA in July 2001, entitled "Preventing Denial of Service Attacks on Network Quality of Service". The paper was published in the proceedings of the conference.
- We presented a paper entitled "ABR Rate Control for Multimedia Traffic Using Microeconomics" at the Intl. Conf. on ATM in June 2000. The paper was published in the proceedings of the conference.

- Douglas Reeves gave a talk entitled "Pricing Goals and Assumptions" at the ENTS workshop on Pricing and QoS in September 1999. There were no proceedings published from this workshop.
- This work was done jointly with NEC C&C Research Labs, Princeton, NJ. Errin and his collaborators obtained a patent on the basic method: "Computer Network with Microeconomic Flow Control", Errin W. Fulp, Maximilian Ott, Daniel Reininger, Patent No. 6,055,571, Dated April 25, 2000

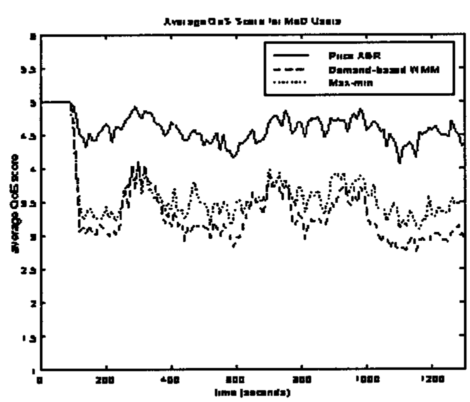
FIGURE I. Experiment showing the superiority of Congestion Pricing over other methods of resource allocation, in terms of its ability to differentiate between different classes of users, and in terms of the net resource utilization.



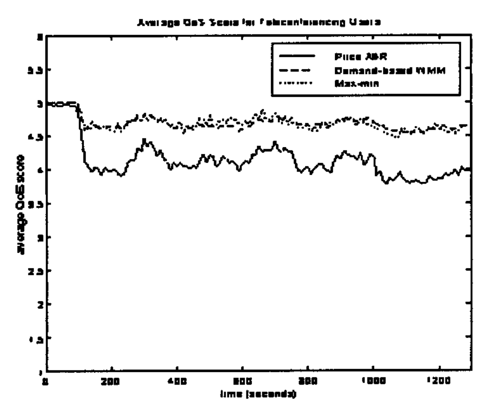
(a) Price ABR link 0 allocation.



(b) Average QoS score for all users.



(c) Average QoS score for MoD users.



(d) Average QoS score for teleconferencing users.

Figure 3: Allocation and average QoS scores.

Topic II: Pricing of Reserved Resources (Risk vs. Reward)

Goal

Many users want the maximum resource amount possible at all times, even if this means their resource allocation will vary over time (in response to demands from other users). Some users value more highly a predictable resource allocation, even if the amount allocated at times may be less than the maximum possible. Users may be modeled as preferring more or less risk, in return for higher or lower average resource allocation. It is desired to have an allocation mechanism that satisfies, or accommodates, both sets of users.

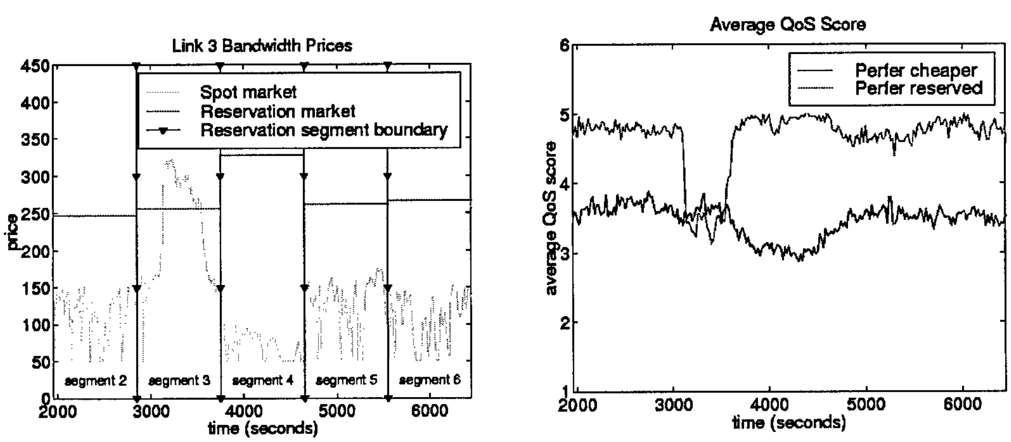
Approach

We implemented two allocation methods, called the "spot" market and the "reservation" market. The reservation market allows users wishing a stable allocation to reserve an amount of resources for a specific length of time. Our mechanism integrates these two mechanisms so that resources are divided between these two groups in a provable optimal way. This is the first work to examine the tradeoff and integrate the two methods.

Results

- We presented a paper at the Networking 2000 in May, 2000, entitled "A Multi-Market Approach to Resource Allocation". This paper was published in the proceedings of the conference.

FIGURE II. Experiment showing how two classes of users ("spot market" = "prefer cheaper", and "reservation market" = "prefer reserved") can both operate in a single market for the resource (spot market prices are more variable, but lower, resulting in a generally higher resource allocation, while the reservation market provides price stability).



Topic III: Protecting Reliable Multicast From Malicious Receivers

Goal

Multicasting is a means of sending data efficiently to a large group of receivers. In a military setting, an example would be distributing commands or intelligence data to a large number of units in the field. A particular type of multicasting, called --reliable multicast--, ensures that all receivers receive exactly the same data, at roughly the same rate, which is critical for some applications (such as the above).

A danger of these approaches is that they rely on requests or feedback from the receivers to determine what the transmission rate should be. One or two receivers, which have been captured or subverted in some way, can falsely claim they cannot receive the data and bring the multicast transmission to a crawl (or a halt) as a result.

Approach

Our solution involves two, distinct components. The first is the development of a distributed method of resource allocation that guarantees the receivers will tell the truth about their actual receiving capability. We discovered that Generalized Vickrey Auctions provide this capability, and we showed how it could be implemented in a distributed fashion.

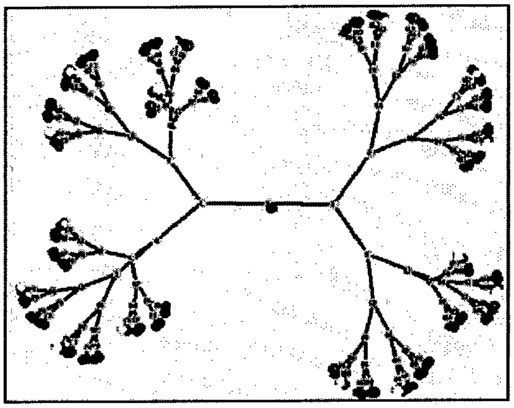
Given this theoretical result, the other challenge was to show how this could be added to an actual reliable multicast method. We chose PGM for this purpose, a well known standard protocol. We demonstrated the vulnerability of PGM to false receiver claims, showed how auction-based allocation could be added to PGM, along with policing of receiver behavior, and then evaluated the effectiveness. A high degree of control of reliability and throughput can be achieved, and misbehaving receivers are detected and blocked from impacting other users.

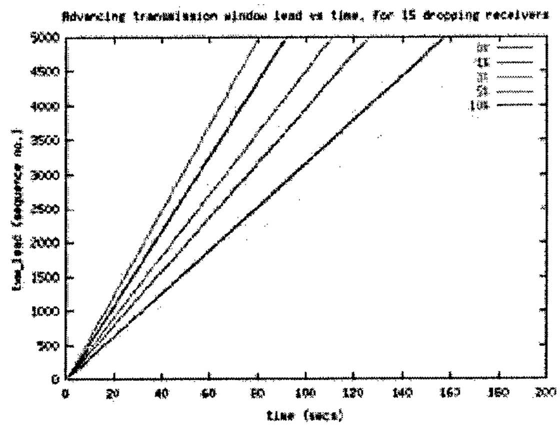
Results

- We collaborated on this work with Prof. Peter Wurman and Ashish Sureka of NC State. They presented a paper on their work entitled "Applying the Generalized Vickrey Auction to Pricing Reliable Multicasts" at the Workshop on Internet Charging and QoS Technology (ICQT 2002) in October 2002. The paper was published in the proceedings of the workshop.
- Nipul Shah defended his MS in Computer Engineering thesis entitled "Preventing Denial of Service Attacks on Reliable Multicast" in December 2002.

FIGURE III. The multicast group on the left has a single sender (red node), approximately 112 "benign" receivers, and 15 "misbehaving" receivers who drop

packets at varying rates. The graph on the right shows how the dropping rate of the misbehaving receivers impacts the data delivery rate for the entire network.





Topic IV: Pricing and Provisioning of Large Scale Resource Amounts

Goal

Allocating resources over the short-term is only part of the decision making to meet user needs. Another critical issue is *provisioning* enough resources over longer periods of time to meet the expected demands from users. This should be done in a way that gives the service provider the most "bang for the buck", that is, for the available resource dollars, meet the needs of the most important and critical applications and users. Knowing how to provision and price resources of differing qualities (and values) is a particularly difficult undertaking.

Approach

Our early pricing work was applied to individual users competing for relatively short-term bandwidth. We expanded our scope to address the decisions of bandwidth providers about long-term bandwidth provisioning, pricing, and allocation to users. We created a model of time-based user demand and showed how to provision to maximize expected profit, and maintain a low blocking (resource non-availability) probability for the users.

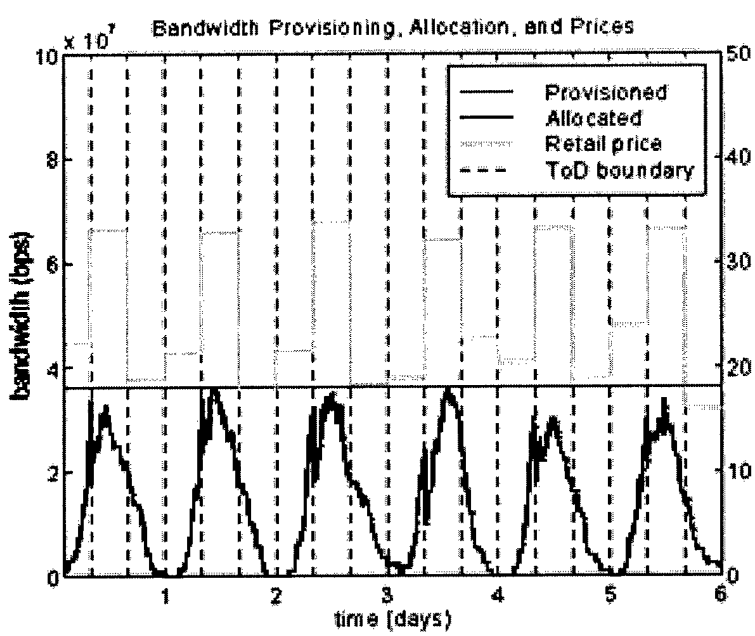
We then applied this work to the case of multiple classes of service, as would be the case in DiffServ (differentiated services) networks run by large organizations. We proposed that user traffic could be dynamically "promoted" to a higher quality of service class if bandwidth was available. Overall user satisfaction is greater and maximum advantage is made of available classes of service.

A key decision in this approach is over what time scale resource prices should remain stable. This is a tradeoff of efficiency (shorter time scales) vs. convenience and predictability for the users (longer time scales). Our investigation indicates the overall gain in value from shorter time scales is quite modest, and probably not justified by the inconvenience to users. Longer timescale allocation is also easier for service providers to implement.

Results

- We presented our work on resource provisioning over longer timescales in a paper entitled "Optimal Provisioning and Pricing of Internet Differentiated Services in Hierarchical Markets" at the Intl. Conf. on Networks (ICN 2001) in June 2001. The paper was published in the proceedings of the conference.
- We presented our work on allocation between different QoS classes in a paper entitled "Optimal Provisioning and Pricing of Internet Differentiated Services Using QoS Class Promotion", at the Workshop on Internet Charging and QoS Technology, in September 2001. The paper was published in the proceedings of the workshop.
- We presented our work on appropriate allocation timescales in a paper entitled "The Economic Impact of Network Pricing Intervals" at the Workshop on Internet Charging and QoS Technology in October 2002. The paper was published in the proceedings of the workshop.

FIGURE IV. Bandwidth is provisioned so that most user needs till be met under normal demands. Demands change frequently, and prices are adjusted at discrete intervals to maximize revenue.



Topic V: Pricing and Provisioning of Different QoS Classes (DiffServ Pricing)

Goal

Differentiated Services (DiffServ) is a standard for providing QoS in the Internet with much less complexity and overhead than other approaches. A key aspect, perhaps *the* key aspect, is provisioning sufficient network resources to meet user expectations about quality. The field of capacity planning is concerned with network design (topology, sizing of links, routing of traffic), but no work has been done on capacity planning for multiple QoS classes.

Approach

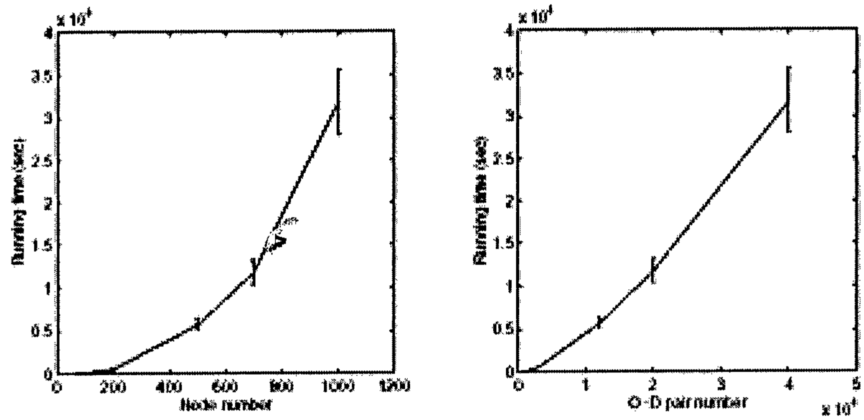
We addressed the problem of capacity planning for two classes of service, Expedited Forwarding (EF) and best-effort (BE). The problem was formulated as a non-linear optimization problem, where the total bandwidth cost was minimized, subject to the QoS requirements of the two classes. Both traffic routing and link capacity assignment were determined, using Langrangian relaxation and subgradient optimization. The solution quality is verifiably good and the running time is very practical for large-sized networks. This represents the first work results on capacity planning of multi-class IP networks with non-linear performance constraints. We believe the method will also be useful for fault-tolerant network design.

Results

We have submitted our work and are awaiting the results:

- "Capacity Planning of DiffServ Networks with Best-Effort and Expedited Forwarding Traffic", submitted to the Intl Conf. on Computer Communications, August 2002.
- "Capacity Planning of DiffServ Networks with Best-Effort and Expedited Forwarding Traffic (extended version)", submitted to Telecommunication Systems Journal.
- "Optimal MPLS Traffic Engineering for Two Classes of Traffic", submitted to the International Teletraffic Congress (ITC-18).

FIGURE V. Demonstration of the running time of our algorithm. Running time grows roughly linearly with the number of users of the network ("O-D pairs"), and is reasonable ($3.6*10^4$ seconds = 10 hours) for very large networks.



Topic VI: Signaling Protocols to Support Resource Pricing

Goal

Congestion pricing is effective at allocating resources fairly and flexibly, with low overhead, when resources are scarce. To make pricing workable, however, there are a number of "infrastructure" issues (accounting, authorization, etc.) that must be resolved, implemented, and deployed. The impact on existing protocols and practices needs to be examined.

Approach

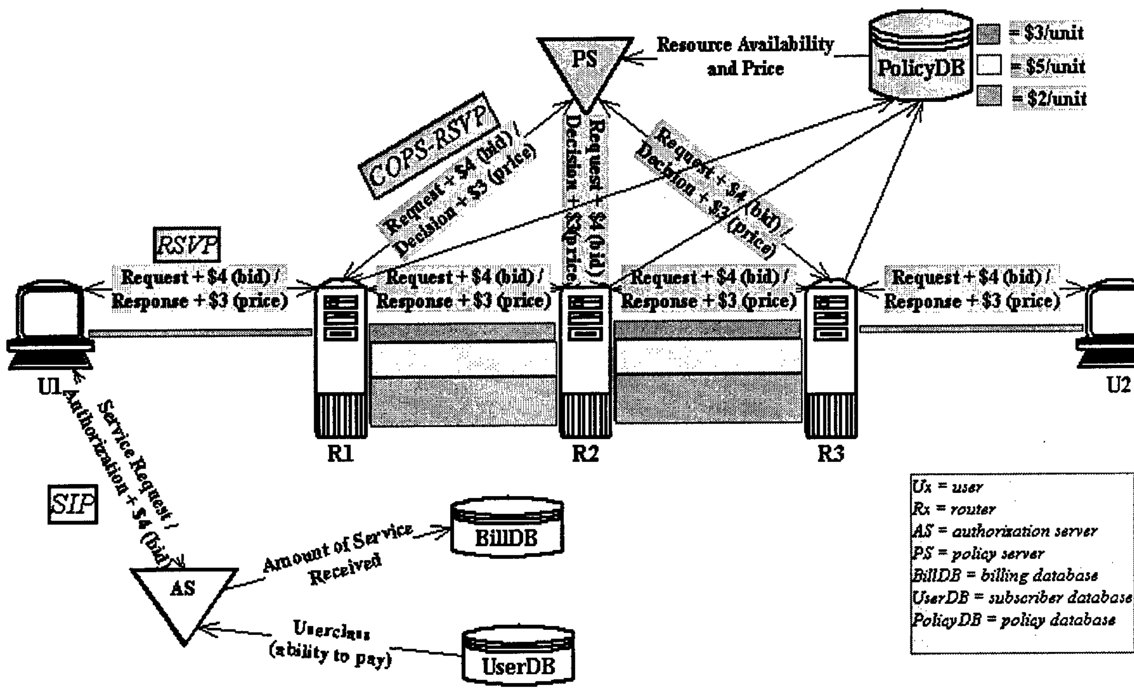
RSVP is an existing resource reservation protocol proposed for use in the Internet. COPS is likewise an existing protocol for communicating resource allocation policies. We have determined and described the changes to COPS and RSVP to convey the information required for pricing of bandwidth. We have implemented these changes, complete with a resource database manager, and authenticated resource pricing requests, in software and demonstrated its capabilities.

Results

- Khurram Khan received his MS in Computer Networking in October of 2002. His thesis title was "COPS Usage for Managing Media Authorization". He is currently employed at Allied Telesyn as a network test engineer.
- Part of this work was done collaboratively with Nortel Networks. As part of that collaboration a patent was filed in October 2000, entitled "A Method for Authorizing Allocation of Resources for Voice Services in IP-Based Networks".
- The software was demonstrated at DISCEX II (Defense Information Security Conference and Expo) in July 2001.

- We are about to submit a standards draft to the Internet Engineering Task Force (IETF) entitled "COPS Usage for Media Authorization".

FIGURE VI. Illustration of the components and messages needed to support pricing of bandwidth with RSVP and COPS protocols.



Summary

For military networks and computers, bandwidth may be scarce or limited during attack, at remote sites, etc. The demands of military priorities require a highly adaptive mechanism for allocating bandwidth that supports a variety of priorties and needs. The resources that are available must also be protected from attack and misuse.

We have made substantial progress towards this goal:

- A method of congestion pricing that is fair, flexible, lightweight, and very adaptable.
- The ability for some users requiring guaranteed resource availability to pay an appropriate "price" for that privilege, in a way that is fair to other users.
- Adaptation of pricing to reliable multicast, to discourage malicious or subverted receivers from degrading the experience of other receivers.
- A method of calculating the resources required to provide satisfactory experience for all users, and to make it profitable for the service provider to do so.

- A way to provision and price bandwidths at different quality levels for different classes of users, and to extract maximum quality from the available resources.
- A demonstration of the software infrastructure (protocols and services) necessary to support pricing of reserved bandwidth for QoS.

Problems Not Addressed

Some problems that were not directly addressed are:

1) Pricing of access to servers, to discourage denial of service attacks on servers.
2) Pricing of media access for wireless networks, to provide fair, flexible allocation of bandwidth.
3) Pricing to control end-to-end delay bounds.

Our reasons for not addressing these problems are several. First, we have come to believe, and others have a similar opinion, that pricing is more appropriate as a mechanism for higher resource amount granularities and larger timescales, so we have concentrated on that. Second, these are closer to specific applications, and we have chosen instead to focus more on fundamental issues than applications. Lastly, for some of these needs other options are now available (for instance, for providing differentiated access to servers). Nevertheless, we believe that the results of this grant have been numerous and substantial, and very much consistent with the goals of our proposal.

Papers Published

- E. Fulp, Z. Fu, D. Reeves, and F. Wu, "Preventing Denial of Service Attacks on Quality of Service", Proc. of DARPA Information Survivability Conference and Exposition (DISCEX II), June 2001.
- E. Fulp and D. Reeves, ``ABR Rate Control for Multimedia Traffic Using Microeconomics", 2nd International Conf. on ATM (ATM '99), Colmar France.
- E. Fulp and D. Reeves, ``QoS Rewards and Risks: A Multi-Market Approach to Resource Allocation", in Lecture Notes in Computer Science, G. Pujolle, ed., No. 1815, pp. 945-956 (Proc. of Networking 2000, Paris, France, May 2000).
- E. Fulp and D. Reeves, "Optimal Provisioning and Pricing of Internet Differentiated Services in Hierarchical Markets", Proc. of Intl. Conf. On Networking, Colmar, France, July 2001.
- E. Fulp and D. Reeves, "Optimal Provisioning and Pricing of Internet Differentiated Services Using QoS Class Promotion", Proc. Of Informatik 2001, Workshop on Advanced Internet Charging and QoS Technology (ICQT 2001).
- Sureka and P. Wurman, "Applying the Generalized Vickrey Auction to Pricing Reliable Multicasts", Proc. of the Workshop on Internet Charging and QoS Technology (ICQT 2002), October 2002.
- E. Fulp and D. Reeves, "The Economic Impact of Network Pricing Intervals", in Proc. of the Workshop on Internet Charging and QoS Technology, October 2002.

Papers Submitted (in review)

- K. Wu and D. Reeves, "Capacity Planning of DiffServ Networks with Best-Effort and Expedited Forwarding Traffic", submitted to the Intl Conf. on Computer Communications.
- K. Wu and D. Reeves, "Capacity Planning of DiffServ Networks with Best-Effort and Expedited Forwarding Traffic (extended version)", submitted to Telecommunication Systems Journal.
- K. Wu and D. Reeves, "Optimal MPLS Traffic Engineering for Two Classes of Traffic", submitted to the International Teletraffic Congress (ITC-18).

Additional Related Activities

The PI worked at Nortel Networks in the Spring and Summer of 2000 on issues related to quality of service in voice-over-IP products.

The PI reviewed several dozen proposals for NSF and AFOSR during this contract period.

The PI presented his work at 3 contractor meetings (at Marina del Ray, Ithaca, and Syracuse) during this period. His group summarized their work to Dr. Herklotz during a visit in September 2000.

The PI is part of the program committees for the 2002 Workshop on Internet Charging and QoS Technology, and DISCEX III (Defense Information Systems Conference and Expo). He has also been invited to participate in a weeklong seminar on Internet Economics in Dagstuhl, Germany in August 2003.

Future Directions

The PI is currently advising 15 graduate students (10 PhD and 5 MS), on problems related to QoS, ad hoc networks, peer-to-peer computing, and network security.